第 18 卷 第 4 期 1975 年 11 月
昆 虫 学 报
ACTA ENTOMOLOGICA SINICA
Vol. 18, No. 4 November, 1975

农作物害虫预测预报的多因子综合相关法

朱 伯 承
(江苏省无锡气象站)

摘要 农作物害虫的发生是由各方面因子决定的，因此在进行害虫预测预报时必须要考虑各方面因子的作用。

本文介绍多因子综合相关的预报方法，把预报量和各预报因子按一定标准化为若干级，用特征资料"0，1"表示之。经直接分析各因子与预报量的"单相关"建立预报方程。

引 言

自然界各种现象，包括害虫的发生，不是孤立存在的，而是存在着相互联系、相互影响、相互制约的关系。例如，害虫的活动，与各种生态因子(气象因子、地理因子、植被群落、其他因子等等)之间存在一定的联系。这种联系如果能用数学关系表示出来，那么我们就可以根据它们之间的数学关系，由预报因子已出现的数值来估计预报对象的数量，从而作出害虫的预测预报。

作者以前介绍的逥归估计法，能建立预报因子与预报量之间的关系(朱伯承，1974)。但由于逥归估计法需解多元线性代数方程，这在预报因子众多的情况下使用起来不很方便。

这里介绍多因子综合相关预报法。本方法将预报量和各预报因子按一定标准分为 n 级，用特征资料"0，1"表示。这种方法最大优点是无需解多元线性代数方程组，经直接分析各因子与预报量的"单相关"即可建立预报方程。这是便于各种害虫预测预报机构使用的方法之一。

基 本 原 理

设有 m 个预报因子 x_1、$x_2\cdots x_m$，要预报的害虫要素用 y 表示。将 x_1、$x_2\cdots x_m$、y 都分为 s 级，则各因子与预报量之间的单相关由列联表(表 1)给出。
并计算单因子相关

$$p^{j}_{l|k}=\frac{n^{j}_{kl}}{n^{j}_{k\cdot}}\qquad (j=1,2\cdots m;\ k,l=1,2\cdots s)\tag{1}$$

将 s 个 s 维单位行向量

$$\begin{pmatrix}1 & 0 & 0 & \cdots & 0\end{pmatrix}$$
$$\begin{pmatrix}0 & 1 & 0 & \cdots & 0\end{pmatrix}$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\begin{pmatrix}0 & 0 & 0 & \cdots & 1\end{pmatrix}$$

表示预报因子 x_j 与预报量 y 的状态，即：

当 y 出现在 1 级时，$y_1=1$，$y_2=0$，$\cdots y_s=0$，

表 1

预报因子 \ 害虫要素		y 1	2	3	…	s	$n_{k\cdot}$
x_j	1	n_{11}^j	n_{12}^j	n_{13}^j	…	n_{1s}^j	$n_{1\cdot}^j$
	2	n_{21}^j	n_{22}^j	n_{23}^j	…	n_{2s}^j	$n_{2\cdot}^j$
	3	n_{31}^j	n_{32}^j	n_{33}^j	…	n_{3s}^j	$n_{3\cdot}^j$
	⋮	…	…	…	…	…	
	s	n_{s1}^j	n_{s2}^j	n_{s3}^j	…	n_{ss}^j	$n_{s\cdot}^j$
$n_{\cdot l}$		$n_{\cdot 1}^j$	$n_{\cdot 2}^j$	$n_{\cdot 3}^j$	…	$n_{\cdot s}^j$	$n_{\cdot\cdot}^j$

当 y 出现在 2 级时，$y_1 = 0$，$y_2 = 1$，$\cdots$ $y_s = 0$，

……………………………

当 y 出现在 s 级时，$y_1 = 0$，$y_2 = 0$，$\cdots$ $y_s = 1$

当 x_j 出现在 1 级时，$x_{j1} = 1$，$x_{j2} = 0$，$\cdots$ $x_{js} = 0$，

当 x_j 出现在 2 级时，$x_{j1} = 0$，$x_{j2} = 1$，$\cdots$ $x_{js} = 0$，

……………………………

当 x_j 出现在 s 级时，$x_{j1} = 0$，$x_{j2} = 0$，$\cdots$ $x_{js} = 1$

第 j 个预报因子对 $y = l$ 级出现的概率贡献为

$$p_l^j = \sum_{k=1}^{s} p_{l|k}^j x_{jk} = p_{l|1}^j x_{j1} + p_{l|2}^j x_{j2} + \cdots + p_{l|s}^j x_{js} \tag{2}$$

对 m 个预报因子，我们取 m 个因子概率贡献的平均值作为 y 为 l 级出现的概率估计值，则有

$$\hat{p}_l = \frac{1}{m} \sum_{j=1}^{m} \sum_{k=1}^{s} p_{l|k}^j x_{jk} \tag{3}$$

由于各项预报因子只能出现在 s 个级别中的一级，因此无论因子出现在哪一级同样恒有

$$\sum_{k=1}^{s} x_{jk} = 1 \tag{4}$$

利用(4)式，可把(3)化为

$$\hat{p}_l = \frac{1}{m} \sum_{j=1}^{m} p_{l|1}^j + \sum_{j=1}^{m} \sum_{k=2}^{s} \frac{1}{m} (p_{l|k}^j - p_{l|1}^j) x_{jk} \tag{5}$$

引进记号

$$\begin{cases} b_{l1} = \dfrac{1}{m} \sum_{j=1}^{m} p_{l|1}^j \\ b_{lk}^j = \dfrac{1}{m} (p_{l|k}^j - p_{l|1}^j) \end{cases} \tag{6}$$

则得预报方程

$$\hat{p}_l = b_{l1} + \sum_{j=1}^{m} \sum_{k=2}^{s} b_{lk}^j x_{jk} \quad (l = 1, 2 \cdots s) \tag{7}$$

这里要指出，$\hat{p}_l$ 是概率值，必须满足

$$\sum_{l=1}^{s} \hat{p}_l = 1 \tag{8}$$

因此，至少并仅需 $s-1$ 个预报方程，第 s 级的预报结果可由 $\hat{p}_s = 1 - \sum_{l=1}^{s-1} \hat{p}_l$ 得到。不过我们不妨用(1)(6)(7)分别建立 s 个预报方程，并由

$$\sum_{l=1}^{s} p^j_{l|k} \equiv 1 \quad (k = 1, 2 \cdots s,\ j = 1, 2 \cdots m) \tag{9}$$

可得

$$\begin{cases} \sum_{l=1}^{s} b_{l1} \equiv 1 \\ \sum_{l=1}^{s} b^j_{lk} \equiv 0 \end{cases} \tag{10}$$

也就是说，(7) 中 s 个预报方程的常数项之和恒等于 1，同一变数项的相应系数之和恒等于 0，这种关系可以帮助我们检验预报方程建立得是否正确。

预报：在作预报时各预报因子出现在哪一级则取该级为 1 其他各级为 0，代入预报方程算出 $\hat{p}$ 各级数值后，即取 $\hat{p}_l$ 值中最大者对应的级别作为预报。

预　报　举　例

下面举无锡地区第一代三化螟蛾发生高峰日期预报的例子。历年（1960—1971 年）无锡县第一代三化螟蛾发生高峰日期（即预报量 y）由表 2 给出。

表　2

年　份	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
y	5/17	5/21	5/26	5/23	5/20	5/30	5/22	5/26	5/27	5/23	5/23	5/27

1.选择预报因子　本例选定如下因子：x_1——无锡县 1 月份雨量，单位毫米；x_2——无锡县 4 月份月平均气压，单位毫巴；x_3——无锡县 4 月份月平均温度，单位度（摄氏）。预报因子的历年值由表 3 给出。

表　3

因子 \ 年份	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
x_1	47.5	42.9	0.2	20.2	67.0	5.5	44.4	8.9	39.0	74.2	15.9	26.4
x_2	1014.8	1014.8	1016.1	1015.5	1012.9	1016.5	1014.7	1016.6	1018.2	1014.8	1017.4	1016.7
x_3	14.1	15.4	13.0	14.3	16.8	12.2	14.0	13.9	14.0	14.5	13.4	13.9

2. 资料处理

（1）预报量 y 分级如下：5 月22日以前为 1 级，5 月 23—25 日为 2 级，5 月 26 日以后为 3 级。

（2）预报因子分级如下：x_1——当 $x_1 \leqslant 10.0$ 毫米时化为 1 级，当 $10.0 < x_1 \leqslant 25.0$

毫米化为 2 级，当 25.0 毫米 $< x_1$ 化为 3 级。x_2——当 $x_2 \leqslant 1014.8$ 毫巴化为 1 级，当 $1014.8 < x_2 \leqslant 1016.0$ 毫巴时化为 2 级，当 1016.0 毫巴 $< x_2$ 化为 3 级。x_3——当 $x_3 \leqslant 14.0$ 度化为 1 级，当 $14.0 < x_3 \leqslant 15.0$ 度化为 2 级，当 15.0度 $< x_3$ 化为 3 级。

预报量与预报因子的分级值由表 4 给出。

表 4

因子 \ 年份	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
x_1	3	3	1	2	3	1	3	1	3	3	2	3
x_2	1	1	3	2	1	3	1	3	3	1	3	3
x_3	2	3	1	2	3	1	1	1	1	2	1	1
y	1	1	3	2	1	3	1	3	3	2	2	3

3. 分析单因子相关 根据表 4 求出列联表（表 5、表 6、表 7）。

表 5

l \ k		y			$n_{k\cdot}$
		1	2	3	
x_1	1	0	0	3	3
	2	0	2	0	2
	3	4	1	2	7
$n_{\cdot l}$		4	3	5	12

表 6

l \ k		y			$n_{k\cdot}$
		1	2	3	
x_2	1	4	1	0	5
	2	0	1	0	1
	3	0	1	5	6
$n_{\cdot l}$		4	3	5	12

表 7

l \ k		y			$n_{k\cdot}$
		1	2	3	
x_3	1	1	1	5	7
	2	1	2	0	3
	3	2	0	0	2
$n_{\cdot l}$		4	3	5	12

并计算各因子的单相关

$$p_{1|1}^1 = 0,\quad p_{2|1}^1 = 0,\quad p_{3|1}^1 = 1,\quad p_{1|2}^1 = 0,\quad p_{2|2}^1 = 1,\quad p_{3|2}^1 = 0,$$

$$p_{1|3}^1 = \frac{4}{7},\quad p_{2|3}^1 = \frac{1}{7},\quad p_{3|3}^1 = \frac{2}{7};$$

$$p_{1|1}^2 = \frac{4}{5},\quad p_{2|1}^2 = \frac{1}{5},\quad p_{3|1}^2 = 0,\quad p_{1|2}^2 = 0,\quad p_{2|2}^2 = 1,\quad p_{3|2}^2 = 0,$$

$$p_{1|3}^2 = 0,\quad p_{2|3}^2 = \frac{1}{6},\quad p_{3|3}^2 = \frac{5}{6};$$

$$p_{1|1}^3 = \frac{1}{7},\quad p_{2|1}^3 = \frac{1}{7},\quad p_{3|1}^3 = \frac{5}{7},\quad p_{1|2}^3 = \frac{1}{3},\quad p_{2|2}^3 = \frac{2}{3},\quad p_{3|2}^3 = 0,$$

$$p_{1|3}^3 = 1,\quad p_{2|3}^3 = 0,\quad p_{3|3}^3 = 0$$

4. 建立预报方程 由(6)式求出

$$b_{11} = 0.3143,\quad b_{21} = 0.1143,\quad b_{31} = 0.5714$$

$$b_{12}^1 = 0,\quad b_{12}^2 = -0.2667,\quad b_{12}^3 = 0.0635$$

$$b_{13}^1 = 0.1905,\quad b_{13}^2 = -0.2667,\quad b_{13}^3 = 0.2857$$

$$b_{22}^1 = 0.3333,\quad b_{22}^2 = 0.2667,\quad b_{22}^3 = 0.1746$$

$$b_{23}^1 = 0.0476,\quad b_{23}^2 = -0.0111,\quad b_{23}^3 = -0.0476$$

$$b_{32}^1 = -0.3333,\quad b_{32}^2 = 0,\quad b_{32}^3 = -0.2381$$

$$b_{33}^1 = -0.2381,\quad b_{33}^2 = 0.2778,\quad b_{33}^3 = -0.2381$$

根据(7)，预报方程为

$$\begin{cases} \hat{p}_1 = 0.3143 + 0.1905x_{13} - 0.2667x_{22} - 0.2667x_{23} + 0.0635x_{32} + 0.2857x_{33} \\ \hat{p}_2 = 0.1143 + 0.3333x_{12} + 0.0476x_{13} + 0.2667x_{22} - 0.0111x_{23} + 0.1746x_{32} - 0.0476x_{33} \\ \hat{p}_3 = 0.5714 - 0.3333x_{12} - 0.2381x_{13} + 0.2778x_{23} - 0.2381x_{32} - 0.2381x_{33} \end{cases} \quad (11)$$

5. 统计历史概括率及预报　在作预报时，将因子按出现在哪一级则取该级为 1，其他级为 0，分别代入预报方程算出 $\hat{p}_l$ 各级数值，按概率最大者对应级别作为预报。

根据所用历史资料对 12 个个例（1960—1971 年）作出的预报列于表 8，与实况对照，12 次中报对 10 次，有 2 次预报错 1 个级别，统计其历史概括率为 83.3%。

表 8

因子 年份	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}	x_{31}	x_{32}	x_{33}	p_1	p_2	p_3	预报	y	检验
1960	0	0	1	1	0	0	0	1	0	0.5683	0.3365	0.0952	1	1	V
1961	0	0	1	1	0	0	0	0	1	0.7905	0.1143	0.0952	1	1	V
1962	1	0	0	0	0	1	1	0	0	0.0476	0.1032	0.8492	3	3	V
1963	0	1	0	0	1	0	0	1	0	0.1111	0.8889	0	2	2	V
1964	0	0	1	1	0	0	0	0	1	0.7905	0.1143	0.0952	1	1	V
1965	1	0	0	0	0	1	1	0	0	0.0476	0.1032	0.8492	3	3	V
1966	0	0	1	1	0	0	1	0	0	0.5048	0.1619	0.3333	1	1	V
1967	1	0	0	0	0	1	1	0	0	0.0476	0.1032	0.8492	3	3	V
1968	0	0	1	0	0	1	1	0	0	0.2381	0.1508	0.6111	3	3	V
1969	0	0	1	1	0	0	0	1	0	0.5683	0.3365	0.0952	1	2	×
1970	0	1	0	0	0	1	1	0	0	0.0476	0.4365	0.5159	3	2	×
1971	0	0	1	0	0	1	1	0	0	0.2381	0.1508	0.6111	3	3	V

我们用预报公式(11)作了 1972、1973、1974 年第一代三化螟蛾发生高峰日期的预报。

1972年：$x_1 = 15.3$ 毫米，$x_2 = 1016.0$ 毫巴，$x_3 = 13.9$ 度。

按因子分级标准化为：$x_{12} = 1$，$x_{11} = x_{13} = 0$；$x_{22} = 1$，$x_{21} = x_{23} = 0$；$x_{31} = 1$，$x_{32} = x_{33} = 0$。代入预报方程算出 $\hat{p}_1 = 0.0476$，$\hat{p}_2 = 0.7143$，$\hat{p}_3 = 0.2381$。其中以 $\hat{p}_2$ 为最大，因此预报 1972 年第一代三化螟蛾发生高峰日期为 5 月 23—25 日，1972 年实际出现的日期为 5 月 25 日。

1973 年：$x_1 = 32.9$ 毫米，$x_2 = 1012.4$ 毫巴，$x_3 = 15.9$ 度。

按因子分级标准化为：$x_{13} = 1$，$x_{11} = x_{12} = 0$；$x_{21} = 1$，$x_{22} = x_{23} = 0$；$x_{33} = 1$，$x_{31} = x_{32} = 0$，代入预报方程算出 $\hat{p}_1 = 0.7905$，$\hat{p}_2 = 0.1143$，$\hat{p}_3 = 0.0952$，其中以 $\hat{p}_1$ 为最大，因此预报 1973 年第一代三化螟蛾发生高峰日期在 5 月 22 日以前，1973 年实际出现的日期为 5 月 19 日。

1974 年：$x_1 = 57.3$ 毫米，$x_2 = 1012.6$ 毫巴，$x_3 = 15.3$ 度。

按因子分级标准化为：$x_{13} = 1$，$x_{11} = x_{12} = 0$，$x_{21} = 1$，$x_{22} = x_{23} = 0$；$x_{33} = 1$，$x_{31} = x_{32} = 0$。代入预报方程可算出 $\hat{p}_1 = 0.7905$，$\hat{p}_2 = 0.1143$，$\hat{p}_3 = 0.0952$。其中以 $\hat{p}_1$ 为最大，故预报 1974 年第一代三化螟蛾发生高峰日期应在 5 月 22 日以前，1974 年实际出现的日期为 5 月 19 日。

从 1972—1974 年的预报看来，三年的级别都预报对了。

讨 论

据近年来多次使用此方法作害虫预测预报,我们有如下一些体会。

1. 本方法的优缺点 优点是:由于原始资料作了(0.1)化处理,又不需求解线性代数方程组,因此计算工作很简单。由于资料分为多级。因而预报的结果也较细,可以分为多种情况作出预报。缺点是:只考虑了每个因子与预报量的相关关系,没有考虑因子之间交叉相关对预报量的贡献。

2. 关于挑选因子 由表 1 可知,若某因子列联表中对角线上的频数大于其所在行的其他各频数,对于因子与预报量正相关(预报举例中因子 2)有

$$n^j_{kk} > n^j_{kl} \text{ 或 } p^j_{k|k} > p^j_{l|k} \quad (k \neq l) \tag{12}$$

对于因子与预报量之间负相关(预报举例中的因子 1、3)有

$$n^j_{k,s-(k-1)} > n^j_{kl} \text{ 或 } p^j_{s-(k-1)|k} > p^j_{l|k} \quad (l \neq s-(k-1)) \tag{13}$$

这样的因子才有预报意义。

另外,若预报因子与预报量独立,则有

$$p^j_{kl} = p^j_{k\cdot}p^j_{\cdot l} \tag{14}$$

对于正相关,因子需符合

$$p^j_{k|k} > p^j_{\cdot k} \tag{15}$$

对于负相关,因子需符合

$$p^j_{s-(k-1)|k} > p^j_{\cdot s-(k-1)} \tag{16}$$

究竟要大多少,因子才是可用的?这要视情况及需要而定。据我们的实践体会,一般说来 20% 左右就可以了。即正相关 $p^j_{k|k} - p^j_{\cdot k} > 20\%$

负相关 $p^j_{s-(k-1)|k} - p^j_{\cdot s-(k-1)} > 20\%$ 的因子是可用的。

在挑选因子时,应尽量挑选因子之间相关小的作为预报因子。若两个完全相关的因子选入,等于其中一个无用。另外,因子的个数不宜太少,否则失去综合考虑的意义,但也不宜太多,太多不仅增加计算量,而且会淹没主要因子的作用。一般说来 3—5 个因子比较适宜。

3. 关于定预报临界值 目前可用两种方法:

(1) 按最高概率原则选定预报临界值(如预报举例中所述)。

(2) 以预报对象出现的平均概率为准,则可规定当计算出的某级概率大于预报对象同级的历史平均概率时即报该级。

若预报对象分为多级,当各级个例数相差不大,而每个级别中有足够多的个例数使计算概率值代表性较好时,可取概率最大值对应的级别作预报。

顺便指出,有时也可按预报要求定临界值。如预报害虫大发生时,希望漏报要少,在定临界值时就可以将预报害虫大发生出现的临界值适当取小些。

参 考 资 料

朱伯承 1974 农作物害虫预测预报的统计学方法。昆虫知识 **11**(2): 39。
复旦大学数学系 1960 概率论与数理统计。上海科学技术出版社。
森口繁一 统计分析。上海科学技术出版社,1961。